Contextual Retrieval in Knowledge Intensive Business Environments

Mark Kröll[1], Andreas S. Rath[1], Michael Granitzer[1], Stefanie Lindstaedt[1], Klaus Tochtermann[1,2]
Know-Center[1] & Knowledge Management Institute TU Graz[2]
A-8010, Graz, Austria
{mkroell, arath, mgrani, slind, ktochter}@know-center.at

Abstract

Knowledge-intensive work plays an increasingly important role in organisations of all types. This work is characterized by a defined input and a defined output but not the way how to transform the input to an output. Within this context, the research project DYONIPOS aims at encouraging the two crucial roles in a knowledge-intensive organization - the process executer and the process engineer. Ad-hoc support will be provided for the knowledge worker by synergizing the development of context sensitive, intelligent, and agile semantic technologies with contextual retrieval. DYONIPOS provides process executers with guidance through business processes and just-in-time resource support based on the current user context, that are the focus of this paper.

1 Introduction

Workflow Management Systems (WFMS) have become quite popular in organizations, because they promise to solve the problems arising from their complex organizational processes. This significant contribution of WFMS to increase the productivity is generally accepted [Riss, 2005].

In spite of the fact that the most important key feature of WFMS that has been identified is the flexibility to deal with changes [van der Aalst *et al.*, 1998; Ellis *et al.*, 1995] van der Aalst and Weske [van der Aalst and Weske, 2005] reference nine articles indicating that Workflow Management Systems are still too restrictive. The usual modeling process where business processes are designed by a process engineer based on interviews and observations of work practices can be seen as a top-down approach. In contrast to this one is the bottom-up approach, referred to as process mining [van der Aalst and Weijters, 2004; Wen *et al.*, 2004; Maruster and Bosch, 2002] where the process model can be derived from workflow, task or/and event logs. In order to enhance the monitored data stored in the logs to tasks or even workflows, innovative information retrieval approaches and mining techniques are needed. The extracted workflow information can be used to model a guide which not just facilitates the work process but also enhances the work quality by just-in-time information retrieval based on the current user, work and organizational context. The consideration of the context in the information retrieval step is on the one hand a big challenge but on the other hand offers the possibility of significant quality improvements [Fuhr, 2005] and ad-hoc accuracy of the results.

This leads us to the objective of this paper which is the presentation of the research project DYONIPOS (DYnamic ONtology based Integrated Process OptimiSation). DYONIPOS strives for encouraging the two crucial roles in a knowledge-intensive organization - the process executer and the process engineer - by synergizing the development of context sensitive, intelligent, and agile semantic technologies with contextual retrieval. The approach of DYONIPOS incorporates the development of solutions based on automatic and semi-automatic knowledge management methods and technologies such as knowledge discovery, semantic systems, and knowledge flow analysis. For a general overview of the DYONIPOS project followed by a detailed description of the advanced features for the process executer and the process engineer we refer to [Tochtermann *et al.*, 2006].

The paper is structured as follows: Section 2 stresses the need for carrying out DYONIPOS and underlines the motivation. The following section outlines the steps DYONIPOS has to pass through in order to provide high quality ad-hoc information to the process executer. In Section 4 is briefly described which algorithms are intended to be used. The paper closes with an overview of the current project state and the list of references.

2 Overview

The goal of DYONIPOS is to solve the dilemma of the organizational need for standardization and control on the one hand and on the other hand the essential freedom of a knowledge worker in his daily job. The research project DYONIPOS aims at mitigating this dilemma by developing context sensitive, intelligent, and agile semantic technologies.

In the business process environment, DYONIPOS will support both the process executer and the process engineer. Process executers are provided with support to find, perform, and record ad hoc processes within their work environment such that the ad hoc process retrieval, application, and definition take place within the executer's current work context. Process engineers will be enabled to review and analyze recorded ad hoc processes, compare them to the standardized processes and automatically enhance them.

The emphasized part of Figure 1 (left part) shows an overview of the DYONIPOS system. It illustrates how DYONIPOS provides information and support for the process executer and the process engineer. The process executer interacts with the system and obtains ad-hoc

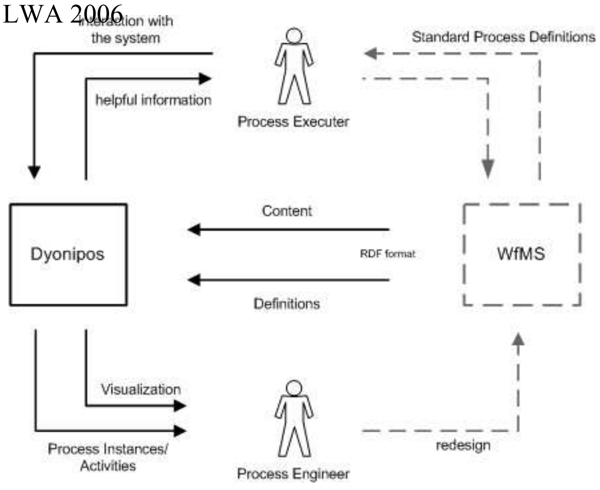

Figure 1: The scope of DYONIPOS

information based on his recent interaction. The process engineer benefits from the preprocessing of identified process structures.

In this paper we mainly focus on the process executer, i.e., featuring guidance through the daily work and providing supportive resources. Resource retrieval is intended to be just-in-time based upon automatic extraction of the context. In the following section we are going to present our ideas and intended procedures.

3 Project Approach

This section is structured in the manner of a question-and-answer game. The question sequence reflects more or less the steps DYONIPOS has to pass through in order to provide the process executer with the promised features. We answer these questions by presenting our approach. To avoid starting each time with "How do we" these three words are replaced by dots.

3.1 ... obtain the context?

First of all we introduce a short synopsis of the required steps to collect information of the process executer's actions and thus to obtain the context [Dey *et al.*, 2001; Coutaz *et al.*, 2005]. It starts with the recording of all events, i.e., the entire user interaction. Events belonging to a logical unit are grouped together into event blocks. Corresponding event blocks form semantic sets and are eventually assigned to the knowledge worker's tasks. Hence one task is represented as a sequence of event blocks. Since event blocks are transfered into graphs, a task can also be modelled as a large graph containing all event blocks in form of subgraphs.

Each level of granularity (events, event blocks and tasks) provides a different representation of the data regarding the semantic quality (see Figure 2). A sequence of events, that is eventually combined to an event block, provides structural information, i.e., how something is done, as well as information due to the content, i.e., what is done. Semantic quality is a measure of this enrichment and is permanently enhanced by passing through the layers and ending at the task level. In the following the individual layers are described in more detail. The data on which DYONIPOS operates consists of the monitored interactions between the user and his computer. Here we use a key logger program, referred to as *event logger*, which records and logs all *events* that occur on the user's computer - a quite similar approach as described by [Fenstermacher, 2005]. Events are user inputs, e.g., mouse movement, mouse clicks, starting a program or creating a folder or file and the reactions to these inputs on the system's side. The sum of all recorded events are stored in the so called *event log*. To ensure security and privacy for the user he has the ability to modify the event log and to delete events from the event log. An important basis for our work in this area build the results from the MISTRAL project [Tochtermann *et al.*, 2005] that aims at extracting semantic concepts from text, audio and video data. It is even conceivable to incorporate into and further process user conversations within the DYONIPOS project.

The knowledge worker is producing quite a bit of data throughout his daily work. Since the event logger is monitoring on a fine granular basis huge amount of data is recorded. To cope with this data mass we perform the following two steps, filtering and relation analysis. Separating relevant from irrelevant data reduces the size of the event log and enhances the output quality for the relation analysis step. In the relation analysis step the events stored in the event log are combined to so called *event blocks* as depicted in Figure 2. Event blocks are built based on predefined static rules. These static rules are a mapping of a set of events to an event block. An example of such a static mapping can be as follows: The user moves the mouse over a program icon, double clicks this icon and the system starts the program. This set of events can be combined to an event block called *starting a program*. An interesting open question here is if it is possible to automatically find a mapping based on the data in the event log, i.e., automatically generating relation rules.

Since not all events and therefore event blocks of a knowledge worker's daily work can be captured automatically the user has the ability to manually add event blocks. Event blocks of this kind can be meeting appointment, talk with a colleague on the corridor, or signing a report.

In Subsection 3.4 it is described how similar event blocks are grouped together into semantic sets. The resulting set represents one task of the process executer. It is intended that the labeling of that task is done automatically by comparing the set of event blocks to other sets. The comparison of entire sets of event blocks is most likely unwise. Thus we select only a few of them (not necessarily sequential) and try to find corresponding event blocks in other sets. Section 4 deals with graph matching algorithms that can be applied to carry out similarity measurements between event blocks.

A high-quality semantic description of the process executer's tasks is thus obtained and is going to be used for further processing. The semantic hierarchy containing three layers is illustrated in Figure 2. Hence, by means of process mining a larger workflow emerges by concatenating individual tasks that were conducted by a number of different knowledge workers.

3.2 ... represent the context?

In DYONIPOS we focus on the knowledge worker's context, i.e., the user context. The user context describes who the user is (organizational context), what he does (work context), how he does it (behavioral context), with whom he is collaborating (social context), and which and what kind of resources he uses (document context). Further contexts that are addressed in DYONIPOS are the process context, that describes the position of the knowledge worker in a business process and the environmental context that captures the nature of the location of the knowledge worker, e.g., computer desktop, meeting room or corridor.

To obtain as much information as possible about the various contexts we rely on recording all user interactions with the system as stated in Subsection 3.1. The collected data is going to be represented as an RDF[1] graph. This representation allows an incorporation and a further processing of relations between graph entities. Certain steps (event blocks) within a task can thus be regarded as graphs where relations between persons, documents and other resources can be easily embedded. These event block graphs are then merged to form a larger graph that represents the entire task and thus part of the user context.

All the different contexts will be related to each other to ensure highly supportive information providing for the knowledge worker. A further application of the contextual information is to identify different and similar tasks. The idea here is to analyze the state of the user context for finding deviations. These deviations could be detectors for switching from one task to another. Identifying context switches could potentially be used as indicators for an update of the provided supportive information. Since contextual retrieval is rather application specific [Fuhr, 2005] further research has to be done in applying contextual retrieval in the area of knowledge-intensive business environments.

3.3 ... store the context?

The knowledge worker's privacy is ensured by law. Thus a natural dilemma arises when trying to gather as much information as possible about the worker's interactions with the system while abiding by the law.

Still, to be able to provide guidance and resources, we have to know about the context of the process executer. As stated in Subsection 3.1, no user interaction remains hidden from the system. Nevertheless data that is going to be stored needs explicit permission. Moreover event blocks are transfered into an abstract form containing the essential data in an encrypted way. The level of encryption is tunable and could be a term vector representation or a hash coding.

3.4 ... exploit the context?

The knowledge worker is not bound to remain within a given task, i.e., executing one task after the other. Switching between tasks might be necessary and even more efficient. In other words, we are given lots of event blocks that have to be grouped together according to their topic as shown in Figure 2. Event blocks that exhibit similar content are identified by analyzing textual information, e.g., which documents were written or read. The degree of similarity indicates the affiliation to a certain set of event blocks. Standard text mining algorithms provide us with the means to extract keywords and compare textual contents.

[1]Resource Description Framework: http://www.w3.org/RDF/

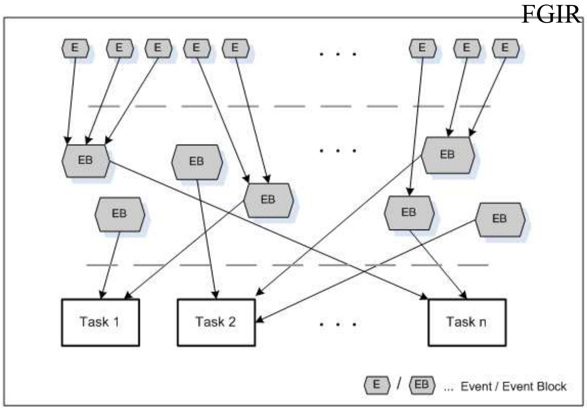

Figure 2: Three semantic layers are shown from the bottom up. The next level can be reached by sensibly grouping together elements from the same layer.

DYONIPOS guides the knowledge worker through given tasks, i.e., suggesting successive actions that have already been performed by other co-workers. Therefore it is necessary to assign the event blocks to the right task. We are going to exploit information regarding the content, e.g., written documents, read reports, visited websites, as well as structural information, e.g., which experts were consulted, which documents were searched for. Similar event blocks of co-workers are identified by applying text mining and graph matching algorithms. Section 4 deals with attempts to exploit the graph structure of event blocks to obtain a measure of similarity. Once similar event blocks are found a possible course of action can be suggested by presenting subsequent event blocks of other process executers.

The guidance for a knowledge worker ends with a recommendation for the next step in the overall process. Therefore the current task has to be assigned to a process, i.e., the position within the overall process has to be figured out. Having assigned the current task, the next step in the process can be identified.

3.5 ... benefit?

The DYONIPOS system provides the process executer with various kinds of resources, e.g., documents other co-workers have read or written performing the same task, useful websites dealing with similar contents, suggesting experts inside or outside the company that could give helpful advice. Resource delivery is based on comparing event blocks as stated in Subsection 3.4. Once similar event blocks of other co-workers have been identified, resources used or generated by them can be proposed. However, in our opinion the presentation of information must not be done prematurely. First the current actions of the knowledge worker have to be assigned to an existing group of event blocks. Resources that have been allocated in the meantime for that group of event blocks are then visualized. If the knowledge worker switches tasks, current resource propositions disappear and DYONIPOS identifies the new content affiliation before providing further resources.

DYONIPOS does not claim to be merely more efficient at retrieving information as traditional search engines. The overall objective is that process executers actually use more information than they would with search engines since there is no effort in obtaining resources. Due to the effort-less accessibility the knowledge worker can incorporate additional resources into his work thus improving the overall work quality.

DYONIPOS intends to follow the same policy as JITIR agents [Rhodes, 2000], i.e., proactivity, presentation of retrieved information in an accessible yet nonintrusive manner, and awareness of the context.

4 Algorithms

As stated above, the internal representation of data will be in form of graphs. In general, given a graph that was constructed by the interaction of the process executer with the system, we would like to find other, similar graphs. In the following it is briefly described which algorithms for measuring the similarity between graphs are intended to be used. Based on the similarity measure, graph classification is carried out. Previous work outlined in [Lux *et al.*, 2006] is utilized for graph retrieval.

Relations between graph entities provide valuable information for mining tasks. Thus, the relatively nascent research areas *Link Mining* [Getoor, 2003] and *Graph Mining* [Washio and Motoda, 2003] successfully exploit the topological view of structured data. Regarding the classification of graphs, relational learning [Mitchell, 1997], finding frequently appearing substructures in graphs [Inokuchi *et al.*, 2000] and kernel methods such as Support Vector Machines [Vapnik, 1995] can be applied. In our work, we focus on kernels for structured data [Gärtner, 2003] and kernel methods [Schölkopf and Smola, 2001].

Convolutional kernels were introduced by [Haussler, 1999] providing a general framework for handling discrete data structures by kernel methods. [Kashima and Inokuchi, 2002; Kashima *et al.*, 2003] concentrated on the construction of graph kernels. Their graph kernel performs a random walk on the vertex product graph of two graphs. The idea behind this kernel is including local information, i.e., taking into account similar edges and vertices of the vicinity.

Another framework of kernel function related with graph structures is called diffusion kernel that was introduced by [Kondor and Lafferty, 2002]. The main difference to the above mentioned approach is that diffusion kernels do not compare two graphs but rather return a similarity measurement between two objects that are represented as vertices of a graph in the input space. Diffusion kernels were applied to document classification [Kandola *et al.*, 2002] where the documents are represented as vertices in a graph.

[Joachims *et al.*, 2001] proposes a combination of kernels each dealing with a different aspect of the data. One kernel deals with the content of objects and another kernel takes the link structure between the objects into account.

Ideas coming outside the world of kernels are considered as well. [Blondel *et al.*, 2004] introduces a concept of similarity between vertices of directed graphs.

5 Project State

The DYONIPOS project is a two year project (March 2006- February 2008) which has started a few months ago. Currently we are implementing semantic technologies and IR methods. The objective is to have a functional environment with which we can match the power of the technologies with our requirements. Furthermore we are analyzing technological synergies with other similar research projects such as AVALON (http://www.iwm.tugraz.at/research/projects/avalon) and MISTRAL (http://www.mistral-project.at).

6 Acknowledgements

The project results have been developed in the DYONIPOS project (DYnamic ONtology based Integrated Process OptimiSation). DYONIPOS is financed by the Austrian Research Promotion Agency (www.ffg.at) within the strategic objective FIT-IT under the project contract number 810804/9338.

The Know-Center is funded by the Austrian Competence Center program Kplus under the auspices of the Austrian Ministry of Transport, Innovation and Technology (http://www.ffg.at), by the State of Styria and by the City of Graz.

References

[Blondel *et al.*, 2004] Vincent D. Blondel, Anahi Gajardo, Maureen Heymans, Pierre Senellart, and Paul Van Dooren. A measure of similarity between graph vertices: Applications to synonym extraction and web searching. *SIAM Rev.*, 46(4):647–666, 2004.

[Coutaz *et al.*, 2005] Joelle Coutaz, James L. Crowley, Simon Dobson, and David Garlan. Context is key. *Commun. ACM*, 48(3):49–53, 2005.

[Dey *et al.*, 2001] A. Dey, D. Salber, and G. Abowd. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications, 2001.

[Ellis *et al.*, 1995] C. Ellis, K. Keddara, and G. Rozenberg. Dynamic change within workflow systems. In *COCS '95: Proceedings of conference on Organizational computing systems*, pages 10–21, 1995.

[Fenstermacher, 2005] Kurt D. Fenstermacher. Revealed processes in knowledge management. In *Wissensmanagement*, pages 397–400, 2005.

[Fuhr, 2005] N. Fuhr. Information retrieval — from information access to contextual retrieval. In M. Eibl, Ch. Wolff, and Ch. Womser-Hacker, editors, *Designing Information Systems. Festschrift für Jürgen Krause*, pages 47–57. UVK Verlagsgesellschaft, 2005.

[Getoor, 2003] L. Getoor. Link mining: a new data mining challenge. *SIGKDD Explor. Newsl.*, 5(1):84–89, 2003.

[Gärtner, 2003] Thomas Gärtner. A survey of kernels for structured data. *SIGKDD Explor. Newsl.*, 5(1):49–58, 2003.

[Haussler, 1999] D. Haussler. Convolution kernels on discrete structures, 1999.

[Inokuchi *et al.*, 2000] Akihiro Inokuchi, Takashi Washio, and Hiroshi Motoda. An apriori-based algorithm for mining frequent substructures from graph data. In *Principles of Data Mining and Knowledge Discovery*, pages 13–23, 2000.

[Joachims *et al.*, 2001] Thorsten Joachims, Nello Cristianini, and John Shawe-Taylor. Composite kernels for hypertext categorisation. In Carla Brodley and Andrea Danyluk, editors, *Proceedings of ICML-01, 18th International Conference on Machine Learning*, 2001.

[Kandola *et al.*, 2002] J. Kandola, J. Shawe-Taylor, and N. Cristianini. On the application of diffusion kernel to text data. Technical report, Neurocolt, 2002.

[Kashima and Inokuchi, 2002] H. Kashima and A. Inokuchi. Kernels for graph classification. ICDM Workshop on Active Mining 2002, 2002.

[Kashima *et al.*, 2003] H. Kashima, K. Tsuda, and A. Inokuchi. Marginalized kernels between labeled graphs. In *Proceedings of the 20th International Conference on Machine Learning*, 2003.

[Kondor and Lafferty, 2002] Kondor and Lafferty. Diffusion kernels on graphs and other discrete structures. In *In Proc. 19th Intl. Conf. on Machine Learning*, 2002.

[Lux *et al.*, 2006] M. Lux, S. Meyer zu Eissen, and M. Granitzer. Graph retrieval with the suffix tree model. In *Proceedings of the Workshop on Text-Based Information Retrieval TIR 06*, 2006.

[Maruster and Bosch, 2002] Laura Maruster and Antal Van Den Bosch. Process mining: Discovering direct successors in process logs. 2002.

[Mitchell, 1997] Thomas M. Mitchell. *Machine Learning*. McGraw-Hill Higher Education, 1997.

[Rhodes, 2000] Bradley J. Rhodes. *Just-In-Time Information Retrieval*. PhD thesis, MIT Media Laboratory, Cambridge, MA, May 2000.

[Riss, 2005] Uwe V. Riss. Knowledge, action, and context: A process view on knowledge management. In *Wissensmanagement*, pages 555–558, 2005.

[Schölkopf and Smola, 2001] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001.

[Tochtermann *et al.*, 2005] K. Tochtermann, M. Granitzer, V. Sabol, and W. Klieber. Mistral: Service-orientierte cross-media techniken zur extraktion von semantik aus multimedia daten und deren anwendung. In *Proceedings Semantics 2005, Reich S., Güntner G., Pellegrini T., Wahler A. (Hrsg.) Trauner Verlag*, 2005.

[Tochtermann *et al.*, 2006] K. Tochtermann, D. Reisinger, M. Granitzer, and S. Lindstaedt. Integrating ad hoc processes and standard processes in public administrations. In *Proceedings of the OCG eGovernment Conference, Linz (Austria)*, 2006.

[van der Aalst and Weijters, 2004] W.M.P. van der Aalst and A.J.M.M. Weijters. Process mining: A research agenda. *Comput. Ind.*, 53(3):231–244, 2004.

[van der Aalst and Weske, 2005] W.M.P. van der Aalst and Mathias Weske. Case handling: A new paradigm for business process support. *Data Knowl. Eng.*, 53(2):129–162, 2005.

[van der Aalst *et al.*, 1998] W.M.P. van der Aalst, G. De Michelis, and C.A. Ellis. Workflow management: Net-based concepts, models, techniques, and tools (wfm'98). In *Computing Science Report 98/7*, 1998.

[Vapnik, 1995] Vladimir N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[Washio and Motoda, 2003] Takashi Washio and Hiroshi Motoda. State of the art of graph-based data mining. *SIGKDD Explor. Newsl.*, 5(1):59–68, 2003.

[Wen *et al.*, 2004] Lijie Wen, Jianmq Wang, Zhe Wang, and Jiaguang Sun. A novel approach for process mining based on event types. 2004.